AF

HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P.O. Box 272400
Fort Collins, Colorado 80527-2400

PATENT APPLICATION

ATTORNEY DOCKET NO. 10992509-2

IN THE
UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s): STINGER

Confirmation No.: 1929

Application No.: 10/693,403

Examiner: Colan, G.

Filing Date: October 24, 2003

Group Art Unit: 2162

Title: AUTOMATICE TABLE DETECTION METHOD AND SYSTEM

Mail Stop Appeal Brief-Patents
Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

TRANSMITTAL OF APPEAL BRIEF

Transmitted herewith is the Appeal Brief in this application with respect to the Notice of Appeal filed on January 18, 2008.

☒ The fee for filing this Appeal Brief is $510.00 (37 CFR 41.20).

☐ No Additional Fee Required.

(complete (a) or (b) as applicable)

The proceedings herein are for a patent application and the provisions of 37 CFR 1.136(a) apply.

☐ (a) Applicant petitions for an extension of time under 37 CFR 1.136 (fees: 37 CFR 1.17(a)-(d)) for the total number of months checked below:

☐ 1st Month $120	☐ 2nd Month $460	☐ 3rd Month $1050	☐ 4th Month $1640

☐ The extension fee has already been filed in this application.

☒ (b) Applicant believes that no extension of time is required. However, this conditional petition is being made to provide for the possibility that applicant has inadvertently overlooked the need for a petition and fee for extension of time.

Please charge to Deposit Account 08-2025 the sum of $ 510. At any time during the pendency of this application, please charge any fees required or credit any over payment to Deposit Account 08-2025 pursuant to 37 CFR 1.25. Additionally please charge any fees to Deposit Account 08-2025 under 37 CFR 1.16 through 1.21 inclusive, and any other sections in Title 37 of the Code of Federal Regulations that may regulate fees.

☒ A duplicate copy of this transmittal letter is enclosed.

☒ I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to:
Commissioner for Patents, Alexandria, VA 22313-1450
Date of Deposit: March 18, 2008

OR

☐ I hereby certify that this paper is being transmitted to the Patent and Trademark Office facsimile number (571)273-8300.
Date of facsimile:
Typed Name: Kelly Davis
Signature:

Respectfully submitted,
STINGER
By

LeRoy D. Maunu
Attorney/Agent for Applicant(s)
Reg No. : 35,274
Date : March 18, 2008
Telephone : 651-686-6633

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appellant:	STINGER	Examiner:	Colan, G.
Serial No.:	10/693,403	Group Art Unit:	2162
Filed:	10/24/2003	Docket No.:	10992509-2 (HPCO.141C1)
Title:	AUTOMATIC TABLE DETECTION METHOD AND SYSTEM		

Board of Patent Appeals and Interferences
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is an Appeal Brief submitted pursuant to 37 C.F.R. § 41.37 for the above-referenced patent application.

I. Real Party in Interest

The real party in interest is Hewlett-Packard Development Company, L.P., having a place of business Houston, Texas. The above referenced patent application is assigned to Hewlett-Packard Development Company, L.P.

II. Related Appeals and Interferences

Appellant is unaware of any related appeals, interferences or judicial proceedings.

III. Status of Claims

Claims 1, 3, 5, 7, 9, 12, and 15-20 are rejected and are presented for appeal. Claims 2, 4, 6, 8, 10, 11, 13, and 14 have been cancelled and withdrawn from consideration. The appealed claims are in the attached Appendix of Appealed Claims.

IV. Status of Amendments

No amendment was filed after final rejection.

V. Summary of Claimed Subject Matter

In the embodiment set forth in claim 1, the invention provides a computer-implemented method of identifying table data in a document. The method includes receiving a page description language representation of the document (FIG. 1, 132; FIG. 2, 202) for providing a list of words in the document and position information for the words (page 9, lines 6-13). The method automatically identifies table data in the document (FIG. 2, 204; FIG. 6, 620; page 10, lines 6-9) based on the page description language representation of the document and at least one table identifying feature. The indentifying includes dividing the document into one or more pages (FIG. 3A 304; page 10, lines 24-29) and dividing each page into a plurality of lines (FIG. 3A 308; page 11, lines 4-7). For each line, the words of the line are clustered into one or more word clusters (FIG. 3A, 312; page 11, lines 8-19; FIG. 6, 640; page 14, lines 25-27). Each cluster includes one or more words, and has a horizontal beginning point, horizontal midpoint, and horizontal end point (page 11, lines 20-22). The identifying further includes comparing alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines (FIG. 3B, #352; page 13, lines 1-14). A cluster in a first line is considered to be aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line (page 13, lines 3-4). The identifying of table data identifies a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table (FIG. 3B, 360, 364; page 13, lines 15-25). The method outputs data descriptive of the lines of the table (FIG. 1, 136, 140, 146; page 7, line 24 - page 8, line 10; FIG. 2, 208; page 9, lines 18-20).

In another embodiment as set forth in claim 7, a computer-readable medium (FIG. 1, 108; page 7, lines 9-18) is provided. The computer-readable medium has stored thereon sequences of instructions, and the sequences of instructions include instructions which, when executed by a processor (FIG. 1, 104), cause the processor to perform the steps including receiving a page description language representation of a document (FIG. 1, 132; FIG. 2, 202) for

providing a list of words in the document and position information for the words (page 9, lines 6-13). The steps also include automatically identifying table data in the document (FIG. 2, 204; FIG. 6, 620; page 10, lines 6-9) based on the page description language representation of the document and at least one table identifying feature. The steps for identifying table data include dividing the document into one or more pages (FIG. 3A 304; page 10, lines 24-29) and dividing each page into a plurality of lines (FIG. 3A 308; page 11, lines 4-7). For each line, the identifying step clusters the words of the line into one or more word clusters (FIG. 3A, 312; page 11, lines 8-19; FIG. 6, 640; page 14, lines 25-27). Each cluster includes one or more words and has a horizontal beginning point, horizontal midpoint, and horizontal end point (page 11, lines 20-22). The alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines is compared (FIG. 3B, #352; page 13, lines 1-14). A cluster in a first line is considered to be aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line (page 13, lines 3-4). The identifying of table data includes identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table (FIG. 3B, 360, 364; page 13, lines 15-25). The method outputs data descriptive of the lines of the table (FIG. 1, 136, 140, 146; page 7, line 24 - page 8, line 10; FIG. 2, 208; page 9, lines 18-20).

The invention as set forth in claim 12 provides a document processing system. The system comprises a processor (FIG. 1, 104; page 7, lines 9-18) for executing programs and a table identification program (FIG. 1, 130; page 7, lines 19-23) for receiving a page description language representation of a document (FIG. 1, 132; FIG. 2, 202). The page description language representation provides a list of words in the document and position information for the words (page 9, lines 6-13). The table identification program automatically identifies table data in the document (FIG. 2, 204; FIG. 6, 620; page 10, lines 6-9) based on the page description representation of the document and at least one table identifying feature. The table identification program is configured to divide the document into one or more pages (FIG. 3A 304; page 10, lines 24-29) and divide

each page into a plurality of lines (FIG. 3A 308; page 11, lines 4-7). For each line, the table identification program clusters the words of the line into one or more word clusters (FIG. 3A, 312; page 11, lines 8-19; FIG. 6, 640; page 14, lines 25-27). Each cluster includes one or more words, each cluster having a horizontal beginning point, horizontal midpoint, and horizontal end point (page 11, lines 20-22). The table identification program compares alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines (FIG. 3B, #352; page 13, lines 1-14). A cluster in a first line is considered to be aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line (page 13, lines 3-4). A line is identified by the table identification program as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table (FIG. 3B, 360, 364; page 13, lines 15-25). The table identification program outputs data descriptive of the lines of the table (FIG. 1, 136, 140, 146; page 7, line 24 - page 8, line 10; FIG. 2, 208; page 9, lines 18-20).

VI. Grounds of Rejection

Claims 1, 3, 5, 7, 9, 12, and 15-20 stand rejected under 35 U.S.C. §102(e) as being anticipated by "Alam" (US Patent 6,336,124 to Alam et al.).

VII. Argument

The rejection of claims 1, 3, 5, 7, 9, 12, and 15-20 should be reversed because the Examiner has not shown that Alam teaches all the limitations of the claims.

Claims 1, 5, 7, 12, and 16

In regards to claim 1, the limitations include "for each line, clustering the words of the line into one or more word clusters, wherein each cluster includes one or more words, each cluster having a horizontal beginning point, horizontal midpoint, and horizontal end point; and for clusters in the plurality of lines,

comparing alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines, wherein a cluster in a first line is aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line; and identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table."

From these limitations it can be seen that the identifying of a line as being part of a table requires the alignment of more than one cluster between a line and a previous line. The Examiner has clearly not shown that Alam teaches these limitations.

Alam generally teaches converting an input document into an intermediate format that is composed of intermediate format blocks, each of which may be a paragraph, a line, a word, or a table (Abstract). The purpose of the intermediate format is to render the document data in a different final output format for display in a chosen different target format (col. 1).

The Examiner cited Alam's col. 12, lines 47-52 and col. 17, lines 10-18 as teaching the limitations of "identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table." The cited portions read are as follows:

> One method of locating tables from a document in the original input format at step 708 generally comprises evaluating a horizontal projection profile of the document, determining upper and lower boundaries of a table by analyzing white space disclosed by the horizontal projection profiles, evaluating a vertical projection profile of the document, and determining a horizontal location of the table by analyzing white space disclosed by the vertical projection profiles. (col. 12, lines 45-52).
>
> ...
>
> FIG. 19 shows a flow diagram of step 1812 for dividing the current block into portions for display such that each portion is within the display parameter or configuration of the display configuration of the output application or device. First, step 1902 determines if the current block is a table. If the current block is not a table, step 1904 breaks up the current block into elements such that each element can be displayed within the display configuration. Each element of a paragraph block may be, for example, a word contained in the paragraph. Other division of a block into elements may be implemented. For example, each element of a list block may be an item or a line in the list. (col. 17, lines 6-18).

From these portions it can be seen that there is no apparent reference to the use of alignment of more than one cluster between a line and a previous line in the identifying of a line as being part of a table. Rather, the column 12 citation deals with locating the boundaries of a table by way of examining projection profiles of the document and analyzing white space, and the col. 17 citation addresses dividing a block into portions for display, and where the block is not a table, breaking up the block into elements that can be displayed. Thus, Alam's col. 12 locates boundaries of a table and does not identify a line as being part of a table based on alignment of clusters. Furthermore, the cited portion of Alam's col. 17 deals with the block not being part of a table. Therefore, the Examiner has not shown that Alam teaches the claimed "identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table."

Elsewhere in col. 17, Alam teaches:

> If the current block is a table, the first row and first column of the table are selected as the row and column headings at step 1905. Although not all first rows and first columns of tables are headings, it can be assumed that the first row and first column are headings. A method may be implemented by which to discriminate between a heading row or column and a data row or column. In addition, some input formats may identify headings of tables and that data can be utilized in this process. (col. 17, lines 27-35).

This portion of Alam does involve processing of table data. However, the processing of the table data is such that the lines of the table have already been identified and grouped into a block. There is no teaching here of how each line is identified as being part of a table, and clearly no suggestion of the claim limitations of the alignment of more than one cluster between lines being used to identify a line of the table.

As further evidence that Alam does not teach the claimed "identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table[,]" the teachings of Alam (col. 10, lines 24-33 and col. 11, lines 4-7) that the Examiner cited as corresponding to the claimed "comparing alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines" do not relate to the identifying a line as being part of a table. The cited teachings of Alam relate to joining lines into paragraphs. A more extensive

quotation from Alam (col. 10, line 6 – col. 11, line 23) is provided below to provide context and for ease of reference.

> FIG. 10 shows a flow diagram illustrating the processing steps for joining the lines into paragraphs after each of the words in the sorted list of words has been assigned to a line.
>
> To join the lines into paragraphs, the first line is assigned to a first paragraph at step 1002. This first paragraph is defined as the current paragraph. A next line is then picked or selected at step 1004.
>
> Preferably, three criteria are met prior to assigning a selected line to a given paragraph. The three criteria are: (1) the selected line is near the paragraph in the Y direction as determined at step 1006; (2) the selected line overlaps the paragraph vertically in the X direction as determined at step 1010; and (3) the words of the selected line have the same font size as the words in the paragraph as determined at step 1012. These criteria and steps 1006, 1010, and 1012 are described in more detail below.
>
> After selecting the next line at step 1004, step 1006 determines whether the selected line is near the current paragraph in the Y direction. To determine whether the selected line is near the current paragraph in the Y direction, the appropriate Y coordinate(s) of the selected line are compared with the appropriate Y coordinate(s) of the previous line of the current paragraph to determine whether certain parameters and/or thresholds are satisfied.
>
> For example, the upper Y coordinate of the selected line may be compared with the lower Y coordinate of the previous line in the current paragraph to determine inter-line spacing in the Y direction. If the inter-line spacing in the Y direction is greater than a threshold, for example, 1.75 times the average character height, then the inter-line spacing threshold in the Y direction is not satisfied and the line is determined not to be near the current paragraph in the Y direction. In addition, if the selected line is at approximately the same position in the Y direction as the previous line in the current paragraph, such as within 10% of the average character height above or below the Y coordinate of the previous line in the current paragraph, the inter-line spacing does not satisfy the minimum inter-line spacing threshold in the Y direction and the line is determined not to be near the current paragraph in the Y direction. Of course, other suitable comparisons and/or analysis may be made by step 1006 to determine whether the selected line is near the current paragraph.
>
> If step 1006 determines that the selected line is not near the current paragraph, step 1008 determines whether the selected line is near any other existing paragraph, i.e., a paragraph which has at least one line assigned thereto. This may be determined with analysis similar to that described above with reference to step 1006.
>
> If step 1006 determines that the selected line is near the current paragraph, or if step 1008 determines that the selected line is near another existing paragraph which is then defined as the current paragraph, step 1010 determines whether the selected line vertically overlaps the current paragraph. A selected line vertically overlaps the current paragraph if the selected line has the same alignment as the current paragraph, for example, left, right or center alignment.
>
> For example, if the left X coordinate of the first word of the current line is within a threshold distance relative to the left X coordinate of the first word of the previous line in the current paragraph, then both the selected line and the current

> paragraph are left aligned and thus overlap. However, as there may be an indented first line in a paragraph, the threshold distance may be defined to be a larger number when comparing the left X coordinate of the first word of the current line with the left X coordinate of the first word of a first line in the current paragraph to account for the hanging indent.
>
> If the right X coordinate of the last word of the current line is within a threshold distance from the right-most X coordinate of the last words of the lines of the current paragraph, then both the selected line and the current paragraph may be right aligned and thus overlap. Further, if the center X coordinate of the current line, i.e., the average of the left X coordinate of the first word and the right X coordinate of the last word of the current line, is within a threshold distance less or greater than the center X coordinate of the previous existing line in the current paragraph, i.e., the average of the left X coordinate of the first word and the right X coordinate of the last word of the previous existing line of the current paragraph, then both the selected line and the current paragraph may be center aligned and thus overlap. The threshold distance may be, for example, 0.5 of the width of a character of the average width of a character.

From the above-quoted portion of Alam it may be observed that this portion does not involve comparing alignment of clusters for use in identifying a line that is part of a table, and that Alam does not compare alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines.

Alam explicitly teaches that FIG. 10 relates to joining lines into paragraphs and that "three criteria are met prior to assigning a selected line to a given paragraph ... : (1) the selected line is near the paragraph in the Y direction as determined at step 1006; (2) the selected line overlaps the paragraph vertically in the X direction as determined at step 1010; and (3) the words of the selected line have the same font size as the words in the paragraph as determined at 1012." Thus, the cited teachings of Alam are clearly not associated with identifying a line as being part of a table.

Appellant further notes that the text quoted above does not teach comparing alignment of horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines. At col. 10, lines 22-33, which the Examiner cited, Alam teaches comparison of Y coordinates between two lines to determine how near one line is to another ("step 1006 determines whether the selected line is near the current paragraph in the Y direction. To determine whether the selected line is near the current paragraph in the Y direction, the appropriate Y coordinate(s) of the selected line are compared with the appropriate Y coordinate(s) of the previous line of the current paragraph to

determine whether certain parameters and/or thresholds are satisfied."). In contrast, the claimed comparing is of alignment of horizontal beginning point, horizontal midpoint, and horizontal endpoint, which is the comparison of X coordinates. Thus, the cited portion of Alam does not suggest these limitations.

Where Alam does look at X coordinates of words on lines, the use does not include comparison of any horizontal midpoint of a word, and that use is not for the purpose of identifying a line as being part of a table. At col. 10, line 55 – col. 11, line 23, Alam's use of X coordinates is for purposes of determining whether the selected line overlaps the paragraph vertically in the X direction. Alam teaches that if the left X coordinate of the first word of the current line is within a threshold distance relative to the left X coordinate of the first word of the previous line then the lines are aligned and overlap and are part of the same paragraph. Alam further teaches that if the right X coordinate of the last word of the current line is within a threshold distance from the right-most X coordinate of the last words of the lines of the current paragraph, then the current line may be right aligned and overlap with the paragraph. For checking center alignment of a line with the paragraph, Alam looks at the average of the left X coordinate of the first word and the right X coordinate of the last word of the current line, relative to the center X coordinate of the previous line in the paragraph. Thus, Alam looks at the left X coordinate of the first word and the right-most X coordinate of the last words on a line to determine the center coordinate of a line. Alam's use of the center coordinate of a line does not correspond to the claimed horizontal midpoint of a cluster. Thus, Alam does not teach comparison of any horizontal midpoint of a word or cluster, and Alam's use of the left and right X coordinates is not for the purpose of identifying a line as being part of a table.

Independent claims 7 and 12 include limitations similar to those of claim 1. Claims 5, 16, and 17 depend from claim 1, and claim 15 depends from claim 12. Therefore, the Examiner has not shown that Alam anticipates these claims for at least the reasons set forth above, and Appellants respectfully request reversal of the rejection.

Claims 3 and 9

According to claim 3, which depends from claim 1, the step of automatically identifying table data in the document based on the number of word

clusters of each line and the alignment of the word clusters between lines includes using the word clusters to generate column position information (the column information includes for each column a horizontal beginning point, horizontal midpoint, and horizontal end point) and updating the column position information by performing a union operation between the column position information of a previous line and the column position information of a current line. The Examiner has not shown that Alam teaches these limitations.

The cited portions of Alam (with additional text quoted for context) are as follows:

> After step 806 determines that the selected word is in the current line or after another existing line is set as the current line at step 809, step 810 determines whether the selected word is within a certain threshold distance or spacing. For example, the appropriate X coordinate of the current selected word is compared with the appropriate X coordinate of the previous word in the current line to determine whether the distance between the words in the X (horizontal) direction are within the threshold distance. In particular, the top left X coordinate of the selected word may be compared with the bottom right X coordinate of the left-most and/or right-most word to determine the spacing between the words in the X direction. If the inter-word spacing in the X direction is greater than a threshold distance, for example, 2.5 times the character width or 2.5 times the average character width, then the inter-word spacing threshold is exceeded and the selected word is determined not to be in the current line. The threshold inter-word spacing in the X direction may be a statistic of the inter-word spacing and may be dynamically determined. Two words positioned approximately at the same vertical position on a page may not be on the same line, for example, when the words are positioned in different columns with spacing between the columns. (col. 8, line 57 – col. 9, line 12).
>
> ...
>
> In one embodiment, improper or erroneous cell breaks between the rows may be determined by locating the upper and lower Y coordinates of each of the rows and determining which of the cell or row breaks may be improper based on the inter-row gaps. For example, the interline spacing within a row may be less than the spacing between two rows. A similar approach may be used to determine improper or erroneous cell breaks between columns. (col. 18, 5-12).

The Examiner further cited teachings from Alam's col. 10, which is quoted above in the arguments for claim 1. From these quoted portions of col. 8 those skilled in the art will recognize that Alam teaches a way to determine whether two words are on the same line. In col. 18, Alam teaches a way to determine improper cell breaks between columns. Alam's col. 10 teachings are inapplicable as explained above in the argument for claim 1. Alam appears to be silent on any use of a horizontal midpoint of a column. Thus, there is no apparent teaching or

suggestion by Alam that the word clusters are used to generate column information that includes a horizontal midpoint, and the Examiner has not shown that Alam teaches all the limitations of claim 3.

Claim 9 depends from independent claim 7 and includes limitations similar to those of claim 3. Therefore, the Examiner has not shown that Alam anticipates claim 9.

Appellant respectfully requests that the rejection of claims 3 and 9 be reversed.

Claims 18, 19, and 20

Claim 18 depends from claim 1 and includes the further limitations of the step of automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters including determining whether the number of word clusters in a line is greater than a threshold value, and classifying the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value. The Examiner has not shown that Alam teaches these limitations.

The cited portion of Alam (with additional text quoted for context) is as follows:

> A determination is made whether the selected word is in the current line at step 806. To determine whether the selected word is in the current line, the appropriate Y coordinate(s), i.e., in the vertical direction, of the selected word are compared with the appropriate Y coordinate(s) of the previous word in the current line to determine whether certain line parameters and/or thresholds are satisfied. For example, the top Y coordinate of the selected word may be compared with the top Y coordinate of the previous word in the current line to determine the inter-word spacing in the Y direction. If the inter-word spacing or distance in the Y direction is greater than a threshold of, for example, 10% of the average character height, then the inter-word spacing parameter in the Y direction is not met and the word is determined not to be in the current line. The average character height may be determined from the words in the current line or from all the words in the document, for example. Of course, other suitable comparisons and/or analysis may be made by step 806 to determine whether the selected word is in the current line. (col. 8, lines 15-34).

From this portion of Alam those skilled in the art will clearly recognize that Alam does not determine whether the number of word clusters in a line is greater than a threshold value. Rather, Alam determines whether a word is in the current line based on the distance between that word and another word in the current line.

There is no apparent suggestion that Alam considers the number of words in a line.

Those skilled in the art will also clearly recognize that Alam does not classify
the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value. The cited portion of Alam is teaching how to determine whether a word is part of a line. In contrast, claim 18 recites how a line is classified as a row in a table (which is inherently after the words have been assigned to a line). Therefore, the Examiner failed to show that Alam teaches the limitations of claim 18.

Claim 19 depends from independent claim 7, and claim 20 depends from independent claim 12. Claims 19 and 20 include limitations similar to those of claim 18. Thus, the Examiner failed to show that claims 19 and 20 are anticipated for at least the reasons set forth above.

Appellant respectfully requests reversal of the rejection of claims 18-20 since the Examiner has not shown that Alam anticipates all the claim limitations.

VIII. Conclusion

In view of the above, Appellant submits that the rejections are improper, the claimed invention is patentable, and that the rejections of claims 1, 3, 5, 7, 9, 12, and 15-20 should be reversed. Appellant respectfully requests reversal of the rejections as applied to the appealed claims and allowance of the entire application.

Respectfully submitted,

CRAWFORD MAUNU PLLC
1270 Northland Drive, Suite 390
Saint Paul, MN 55120
(651) 686-6633

By: 

Name: LeRoy D. Maunu
Reg. No.: 35,274

APPENDIX OF APPEALED CLAIMS FOR APPLICATION NO. 10/693,403

1. A computer-implemented method of identifying table data in a document comprising the steps of:

receiving a page description language representation of the document for providing a list of words in the document and position information for the words; and

automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature, wherein the identifying step includes,

dividing the document into one or more pages;

dividing each page into a plurality of lines;

for each line, clustering the words of the line into one or more word clusters, wherein each cluster includes one or more words, each cluster having a horizontal beginning point, horizontal midpoint, and horizontal end point;

for clusters in the plurality of lines, comparing alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines, wherein a cluster in a first line is aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line; and

identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table; and

outputting data descriptive of the lines of the table.

3. The method of Claim 1 wherein the step of automatically identifying table data in the document based on the number of word clusters of each line and the alignment of the word clusters between lines further comprises:

using the word clusters to generate column position information, wherein the column information includes for each column a horizontal beginning point, horizontal midpoint, and horizontal end point; and

updating the column position information by performing a union operation between the column position information of a previous line and the column position information of a current line.

5. The method of Claim 1 wherein receiving a page description language representation of the document for providing a list of words in the document and position information for the words includes receiving a PDF representation of the document, and wherein converting the table data encompassed by each table bounding box to a markup language representation includes converting the table data encompassed by each table bounding box to a HTML representation.

7. A computer-readable medium having stored thereon sequences of instructions, said sequences of instructions including instructions which, when executed by a processor, cause said processor to perform the steps of:

receiving a page description language representation of a document for providing a list of words in the document and position information for the words; and

automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature, wherein identifying includes,

dividing the document into one or more pages;

dividing each page into a plurality of lines;

for each line, clustering the words of the line into one or more word clusters, wherein each cluster includes one or more words, each cluster having a horizontal beginning point, horizontal midpoint, and horizontal end point; and

for clusters in the plurality of lines, comparing alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines, wherein a cluster in a first line is aligned with a cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is

aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line; and

identifying a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table; and

outputting data descriptive of the lines of the table.

9. The computer-readable medium of Claim 7 further containing instructions which, when executed by said processor, would cause said processor to perform the steps of:

using the word clusters to generate column position information, wherein the column information includes for each column a horizontal beginning point, horizontal midpoint, and horizontal end point; and

updating the column position information by performing a union operation between the column position information of a previous line and the column position information of a current line.

12. A document processing system comprising:

a processor for executing programs; and

a table identification program for receiving a page description language representation of a document, the page description language representation providing a list of words in the document and position information for the words, and for automatically identifying table data in the document based on the page description representation of the document and at least one table identifying feature, wherein the identification program is configured to,

divide the document into one or more pages;

divide each page into a plurality of lines;

for each line, cluster the words of the line into one or more word clusters, wherein each cluster includes one or more words, each cluster having a horizontal beginning point, horizontal midpoint, and horizontal end point;

for clusters in the plurality of lines, compare alignment of the horizontal beginning point, horizontal midpoint, and horizontal end point of clusters between lines, wherein a cluster in a first line is aligned with a

cluster in a previous line if at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the first line is aligned with at least one of the horizontal beginning point, horizontal midpoint, and horizontal end point of the cluster in the previous line; and

identify a line as being part of a table in response to more than one cluster of the line being aligned with clusters of previous lines identified as part of the table; and

output data descriptive of the lines of the table.

15. The document processing system of claim 12 wherein the table identification program further comprises:

a conversion module coupled to the bounding box generation module for receiving the table bounding box for each table in the document, and for converting the words encompassed by the table bounding box into a markup language representation that maintains the table structure of each table.

16. The method of claim 1 wherein the step of automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature further comprises:

automatically identifying table data in the document based on one or more table headings.

17. The method of claim 1 wherein the step of automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature further comprises:

automatically identifying table data in the document based on one or more horizontal lines and vertical lines that separate rows or columns of the table.

18. The method of claim 1, wherein the step of automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters comprises:

determining whether the number of word clusters in a line is greater than a threshold value; and

classifying the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value.

19. The computer-readable medium of claim 7, wherein the instructions for automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters include instructions that when executed by a processor cause the processor to perform the steps further comprising:

determining whether the number of word clusters in a line is greater than a threshold value; and

classifying the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value.

20. The document processing system of claim 12, wherein the table identification program is further configured to:

determine whether the number of word clusters in a line is greater than a threshold value; and

classify the word clusters in the line as a row of a table in response to the number of word clusters in a line being greater than the threshold value.

APPENDIX OF EVIDENCE FOR APPLICATION NO. 10/693,403

Appellant is unaware of any evidence submitted in this application pursuant to 37 C.F.R. §§ 1.130, 1.131, and 1.132.

APPENDIX OF RELATED PROCEEDINGS FOR APPLICATION NO. 10/693,403

Appellant is unaware of any related appeals, interferences or judicial proceedings.